Current Biology, Vol. 14, 150–154, January 20, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.cub.2003.12.055

Selection on Codon Usage in *Drosophila americana*

Xulio Maside,* Angela Weishan Lee, and Brian Charlesworth
Institute of Cell, Animal and Population Biology
University of Edinburgh
Kings Buildings
West Mains Road
EH9 3JT Edinburgh
UK

Summary

Synonymous codons are not used at random, significantly influencing the base composition of the genome. The selection-mutation-drift model proposes that this bias reflects natural selection in favor of a subset of preferred codons [1–4]. Previous estimates in *Drosophila* of the intensity of selective forces involved seem too large to be reconciled with theoretical predictions of the level of codon bias [4]. This probably results from confounding effects of the demographic histories of the species concerned. We have studied three species of the *virilis* group of *Drosophila*, which are more likely to satisfy the assumptions of the evolutionary models. We analyzed the patterns of polymorphism and divergence in a sample of 18 genes and applied a new method for estimating the intensity of selection on synonymous mutations based on the frequencies of unpreferred mutations among polymorphic sites. This yielded estimates of selection intensities ($N_e s$) of the order of 0.65, which is more compatible with the observed levels of codon bias. Our results support the action of both selection and mutational bias on codon usage bias and suggest that codon usage and genome base composition in the *D. americana* lineage are in approximate equilibrium. Biased gene conversion may also contribute to the observed patterns.

Results and Discussion

Many species exhibit nonrandom usage of alternative codons for the same amino acid [5]. Correlations of the codon usage bias of genes with features such as tRNA abundance [6] and level of gene expression [7] suggest that some species experience natural selection for "preferred" codons. The intensity of such selection has been estimated by fitting population genetics models to the observed frequency distributions of synonymous variants in natural populations of *Drosophila* [3, 8]. The results suggest that $N_e s$ is typically of the order of 1–2, where N_e is the effective population size of the species, and s is the selection coefficient against an unpreferred codon at a given site [3, 8]. Given estimates of N_e of 10^6 or more, s must be of the order of 10^{-6} [3, 8]. There are, however, several difficulties with these results. The base composition of some species is apparently not in equilibrium such that unpreferred codons are accumulating [9, 10]. One species appears to have undergone a recent population expansion [11], which is likely to distort the frequency distributions of synonymous variants. Finally, an $N_e s$ value of 1 or more means that genes should use nearly 100% preferred codons in contrast to what is observed [4]. The current estimates of $N_e s$ for codon preferences in *Drosophila* must therefore be regarded with skepticism.

Here, we have used data on DNA sequence polymorphisms in *D. americana* and sequence comparisons with its close relatives *D. virilis* and *D. ezoana* to reexamine the problem of estimating selection intensities. We have applied a new method of estimation that avoids some of the difficulties just described. *D. americana* is a member of the *virilis* subgroup of *Drosophila* and is more closely related to *D. virilis* (which diverged less than 10 million years ago [mya]) than to *D. ezoana* (11 to 15 mya) [12]. Its ecology means that it is less likely to have been disturbed by human activities than species such as *D. melanogaster* and *D. simulans* [13]. This set of species is thus suitable material for between-species comparisons and population genetic studies.

The action of selection can be detected from the ratio of polymorphism to divergence for synonymous mutations [3, 8]. Selection is less effective at preventing the establishment of weakly deleterious mutations than at preventing their fixation. The ratio of the number of deleterious mutations from preferred to unpreferred codons (P→U) relative to advantageous mutations in the opposite direction (U→P) will be higher among polymorphic variants within a species than among mutations that distinguish related species [3, 8]. To test for this effect, we studied codon usage and patterns of within and between-species variation at 18 nuclear genes (see Table S1). Given previous evidence for shifts in codon preferences between *D. virilis* and *D. melanogaster* [14], we first constructed a table of preferences for *D. virilis* (see Experimental Procedures). Levels of codon usage bias were measured by the frequencies of optimal codons in each gene (*Fop*, [15], Table 1). Variants were classified as either polymorphisms within *D. americana* or as fixed differences between *D. virilis* and *D. americana*. Their ancestral state (P versus U) was inferred by parsimony [16], using *D. ezoana* as the outgroup.

The fraction of P→U changes among polymorphic mutations (Table 2) is significantly larger than for fixed differences (the ratios of polymorphic to fixed differences, r_{pd}, are 2.38 and 0.64 for P→U and U→P, respectively; $p < 0.001$ from Fisher's exact test), consistent with the action of selection. Given that a fraction of the changes segregating in the population are misclassified as fixed, because of limited sample sizes, these are probably underestimates of the true ratios. After partially correcting for this bias (see Experimental Procedures), the two ratios are 3.07 and 0.70. Importantly, the corrected numbers of P→U and U→P fixed substitutions are not signifi-

*Correspondence: xulio.maside@ed.ac.uk

Table 1. Codon Usage Bias (*Fop*) in *D. americana*

Gene	Chromosome	*Fop*
su(Hw)	2	0.75
anon-66Db	3	0.59
cdc37	3	0.61
ddx1	3	0.57
dos	3	0.69
kni	3	0.68
msl3	3	0.59
rh4	3	0.65
sina	3	0.66
tll	6	0.77
cps36	X	0.73
csw	X	0.72
elav	X	0.77
fused1	X	0.60
per	X	0.72
pros28.1	X	0.45
su(s)	X	0.79
yp1	X	0.66

cantly different ($p > 0.16$; χ^2 for 1:1 ratio), as expected if codon usage is in equilibrium in these species [3].

In *D. virilis*, as in many other organisms, preferred codons usually end in G or C (see Experimental Procedures). Therefore, nonselective factors such as recent changes in the magnitude of mutational bias [17] and/or GC-biased gene conversion [18, 19] can have similar effects to those of selection. Given that these processes should affect exons, introns, and intergenic sequences in a similar way in a given region of the genome [20], their prevalence can be tested by comparing the ratio of r_{pd} (GC→AT) to r_{pd} (AT→GC) in coding (r_c) versus noncoding (r_{nc}) DNA. Selection on codon usage predicts that $r_c > r_{nc}$, whereas the other processes would be expected to cause similar substitution patterns in both types of sequences ($r_c = r_{nc}$). In our data set, r_c is much larger than r_{nc} (Table 3). This difference is caused by a much larger fraction of polymorphic GC→AT changes in exons than introns ($p < 0.001$; Fisher's exact test), reflecting their higher GC content (means of 0.62 versus 0.39). On the other hand, the number of fixed changes in either direction does not differ from 1:1 in either exons or noncoding sequences ($p > 0.20$ and $p > 0.50$, respectively; χ^2 test). This could also be explained by a lack of BGC in noncoding DNA due to much lower recombination rates in these regions. But we found that the recombination rate per site does not differ between introns and exons in *elav* (maximum likelihood estimates of $4N_e r = 0.086$ and 0.092, respectively, where r is the rate of recombination between adjacent base pairs; see Figure S1), the one gene for which we have sufficient polymorphism data to estimate recombination rates.

Another test for natural selection is to compare the frequency distributions of polymorphic P→U and U→P mutations [21]. Under the mutation-selection-drift model, deleterious P→U changes should segregate at lower frequencies than advantageous U→P mutations. Consistent with this, we observed that P→U polymorphisms in *D. americana* are found at significantly lower frequencies than U→P mutations (Figure 1; Mann-Whitney test $U_s = 2.23$, $p < 0.05$). In addition, we compared the frequency spectrum of P→U and U→P mutations with neutral expectation by means of Fu and Li's *D* test for the numbers of mutations which are represented only once in the sample [22]. Fu and Li's *D* statistic for P→U changes was negative in 13 out of the 18 genes, reaching significant values in the cases of *cdc37* and *elav* (Table S2); there was also a significant excess of P→U singletons in the pooled sample ($D = -3.80$; $p < 0.001$). In contrast, U→P substitutions and mutations in noncoding sequences did not depart significantly from neutral expectations. Our data thus support the action of selection on codon usage bias and suggest that the codon usage and genome base composition in the *D. americana* lineage are in approximate equilibrium.

We have also applied a new method for estimating the intensity of selection at synonymous sites based on

Table 2. Polymorphism and Divergence of Synonymous SNA Changes in *D. americana*

			Polymorphic		Fixed		
Gene	Chromosome	N	P→U	U→P	P→U	U→P	Fisher's Exact Test
su(Hw)	2	5	8	1	2	2	
anon-66Db	3	5	12	0	5	2	
cdc37	3	5	4	1	2	1	
ddx1	3	6	8	2	5	1	
dos	3	6	12	5	2	2	
kni	3	5	3	0	4	4	
msl3	3	5	2	1	1	0	
rh4	3	5	7	1	2	3	
sina	3	5	5	1	1	0	
tll	6	8	21	3	5	2	
cps36	X	5	3	0	4	0	
csw	X	5	2	2	1	2	
elav	X	50	16	1	2	2	
fused1	X	5	1	0	1	1	
per	X	5	5	0	6	2	
pros28.1	X	52	5	1	0	1	
su(s)	X	5	7	0	4	4	
yp1	X	5	3	1	2	3	
Total			124	20	52	32	$p < 0.001$
Corrected			133	21	43	31	$p < 0.001$

Table 3. Polymorphism and Divergence of GC→AT and AT→GC Changes at Coding and Noncoding Sites in *D. americana*

	GC→AT	AT→GC		Fisher's Exact Test
Coding DNA, polymorphic	145	28		
Coding DNA, fixed	43	32		$p < 0.001$
	$r_{pd} = 3.4$	$r_{pd} = 0.9$	$r_c = 3.8$	
Noncoding DNA, polymorphic	20	20		
Noncoding DNA, fixed	9	7		$p = 0.77$
	2.1	2.7	$r_{nc} = 0.8$	

These values have been corrected to account for the fraction of polymorphisms misclassified as fixations (see Experimental Procedures). Data on noncoding DNA has been obtained from the ~2.2 kb of noncoding sequences represented in our data set (see Table S1).

the frequencies of P→U mutations among sites polymorphic for P→U or U→P mutations (see Experimental Procedures). The scaled selection parameter $4N_es$ [3, 8] is denoted by γ (this applies to both autosomal and sex-linked loci, although the value of N_e may differ according to mode of inheritance [4]). We allow for mutational bias by assuming that the P→U mutation rate is k times the U→P mutation rate. This yields a likelihood equation for the data (equation 3) on the assumption of independence between different sites and statistical equilibrium under mutation, selection, and drift.

The genes for which polymorphism data were available were divided into two classes with low and high bias, respectively, each with nine genes. Equation 3 was applied to each class and the log-likelihood curves examined as a function of γ. In addition, the whole set of genes was examined for the fit of a common γ value. For low-bias genes, the maximum likelihood was at $\gamma = 2.6$ ($-\ln L = -23.49$; two-unit support limits at 1.6 and 3.7) and for high-bias genes at 2.7 ($-\ln L = -35.33$; two-unit support limits at 1.9 and 3.6). The pooled set had a maximum at $\gamma = 2.6$ ($-\ln L = -58.84$; two-unit support limits at 2.0 and 3.5). This differs from the sum of the ln L values for the two classes by only 0.02, so there is no statistically significant evidence for differences in the strength of selection among classes.

This is interesting in view of the fact that there is a positive correlation between the GC content of third codon positions (*GC3*) and of noncoding sequences (*GCnc*) ($r^2 = 0.34$; $p < 0.05$, on a Kendall's 2-tailed test). A similar correlation was also observed in a larger sample of 59 *D. virilis* genes for which there are exon and intron data. These results suggest that the differences in codon usage between the genes may to some extent reflect differences in mutational bias as well as the strength of selection. It worth noting, however, that this association seems to be restricted to the low-bias genes (Figure 2), as would be expected if selection was more relaxed in this subset of genes [23].

Our pooled estimate of N_es is 0.65, which is much smaller than estimates derived by the frequency spectrum methods for *D. simulans* and *D. pseudoobscura* [24, 8] and is compatible with the presence of unpreferred codons segregating in the population [4]. Interestingly, maximum likelihood fits of the data to the frequencies of U singletons in our samples yield γ values of 5.3 and 3.5 for low- and high-bias genes, respectively, which may reflect a recent population expansion in *D. americana*. Theoretical work (A. Veron and B.C., unpublished data) indicates that short-term changes in population size have much smaller effects on the proportion of P→U polymorphisms than on frequency spectra.

Given that most preferred codons in this species end in G or C, *Fop* and *GC3* are closely related. It is, however,

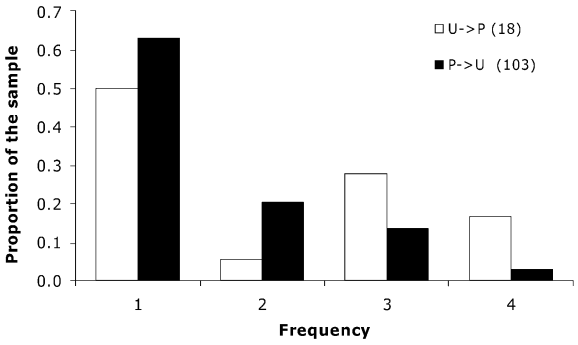

Figure 1. Frequency Distribution of Synonymous Variants in *D. americana*

The frequencies were calculated from a sample of five alleles for each gene.

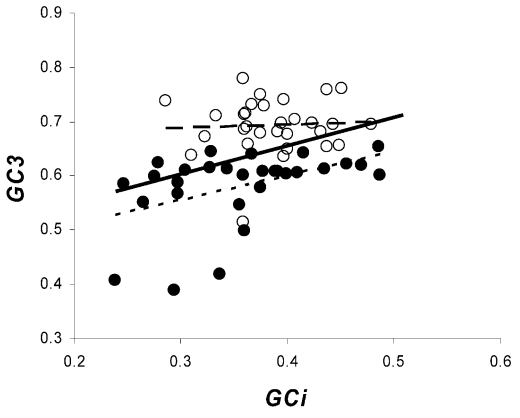

Figure 2. Correlation between GC Content at Synonymous (*CG3*) and Intronic Sites (*GCi*) in *D. virilis*

The sample consists of 59 full-length, nonredundant genes with at least 100 bp of intron sequence, available in GenBank. *GC3* and *GCi* are positively correlated in the overall sample ($r^2 = 0.16$, $p < 0.01$, Kendall's two-tailed test, continuous regression line). To test if this effect is independent of the intensity of selection on codon usage, we divided the sample in two equal size ($n = 30$ and 29) groups comprising high- (*Fop* ≥ 0.65) and low- (*Fop* < 0.65) bias genes, respectively. Regression analyses revealed a similar significant association among low-bias genes ($r^2 = 0.20$, $p < 0.05$; black circles, dotted line), but not among the high-bias genes ($r^2 = 0.00$, $p > 0.95$; closed circles, segmented line).

difficult to account for the magnitude of these parameters solely by selection and mutational bias, as can be seen as follows. Let k' and γ' be the mutation bias and the scaled selection parameter ($4N_e s$) for silent GC ↔ AT changes, respectively. The mean *GCnc* of our set of genes is about 0.39. If this reflected neutral equilibrium, the corresponding value of k' from equation 2 would be 1.6 [1, 4]. About 20% of all possible GC ↔ AT synonymous changes in our genes imply a P→P or U→U change, which are likely to be effectively neutral, so that γ' can be approximated as 0.8 γ. The mean *GC3* is 0.66, but with $k' = 1.6$ and $\gamma' = 1.6$ (using $\gamma = 2.0$, the lower support limit for the complete set of genes), the predicted *GC3* is 0.76. One possibility is that BGC affects both coding and noncoding sequences in opposition to mutational bias in favor of AT, but we have failed to detect its effects because of the small number of noncoding polymorphisms (Table 3). This would imply a nonzero value of γ' for noncoding sequences and, therefore, a larger value of k', with a consequent reduction in the expected *GC3*.

To explore this, we applied equation 3 to the pooled sample of noncoding polymorphisms. This gave a maximum-likelihood estimate of $\gamma' = 0$, with an upper two-unit support limit of approximately 0.8. A *GCnc* content of 0.35, typical of the low-bias genes, would be produced with $k' = 3.1$ and $\gamma' = 0.5$. *GC3* for this set of genes is 0.60, with $k' = 3.1$; this corresponds to $\gamma' = 1.5$, which is within the support limits for coding sequences of low-bias genes (these limits can be approximated as 0.8 times the limits for γ given above). Similarly, with $\gamma' = 0.5$ for the noncoding regions, the mean *GCnc* of 0.44 for high-bias genes requires $k' = 2.1$; their mean *GC3* of 0.72 is consistent with $\gamma' = 1.7$, within the two-unit support limits for coding sequences of high-biased genes. These data are thus consistent with the hypothesis that both BGC and mutational bias affect the GC content of coding and noncoding regions. Further data on r_{nc} are needed to test this hypothesis.

Experimental Procedures

Codon Usage in the *virilis* Group

The preferred codons in *D. virilis* were identified by means of a correspondence analysis of codon usage [25] in the 129 *D. virilis* nuclear genes available in GenBank (incomplete CDSs and/or redundant sequences were not included in the data set) using the program CodonW [26] available at http://www.molbiol.ox.ac.uk/cu/codonW.html. The resulting set of preferred codons is slightly different from that proposed for *D. melanogaster* [16]: in the case of Tyr, His, and Asp, all of them 2-fold degenerate amino acids, the U ending codons were used more frequently than their C ending synonyms. These differences were not significant and therefore no codon was considered as preferred for these aminoacids. For Ser, Pro, Thr, and Ala it was possible to identify one additional preferred codon in comparison with those found in *D. melanogaster* (AGC, CCG, ACG, and GCG, respectively). Given the reduced low level of synonymous divergence among them, we assumed in this study that the preferences in *D. virilis* are conserved in *D. americana*.

Correction for Polymorphic Changes Misclassified as Fixations

Under an equilibrium neutral model, the proportion of segregating sites that are detected as polymorphisms in a sample of n alleles is approximately $s_{n-1}/\ln(2N)$, where $s_{n-1} = 1 + (1/2) + (1/3) + \ldots + 1/(n-1)$, and N is the population size ([27], p. 276). The chance that a mutant that is segregating at frequency x in the population is fixed in a sample of n alleles is x^n. Since the probability that a mutant is found at frequency x is approximately $x^{-1}/\ln(2N)$ ([27], p. 276], the net probability that a polymorphic mutant is misclassified as fixed is approximately equal to the integral between 0 and 1 of $x^{n-1}/\ln(2N)$, i.e., $n^{-1}/\ln(2N)$. If the observed number of neutral polymorphisms in the population in a genomic region is S_p, the estimate of the number that are misclassified as fixed is thus $S_p(n\, s_{n-1})^{-1}$.

Analysis of Data on Codon Usage and Synonymous Polymorphisms

We assume additive selection such that homozygotes and heterozygotes for the unpreferred codon at a given position have fitnesses reduced to $1 - 2s$ and $1 - s$, respectively, relative to the fitness of homozygotes for the preferred codon. We consider only polymorphisms that involve P→U or U→P mutations, where P denotes a preferred codon and U an unpreferred codon.

Assume that the sojourn time densities ([27], pp. 148–149]) at allele frequency x in the interval $1/(2N)$ to $1 - 1/(2N)$ are $\phi_0(x)$ for P→U mutations, and $\phi_1(x)$ for U→P mutations, respectively. Assume that the P→U mutation rate is k times the U→P mutation rate. For a sample of n alleles from the population, standard results ([27], pp. 275–277) imply that the probability of observing a P→U mutation at a segregating site is $\pi = kpI_0/(kpI_0 + [1-p]\, I_1)$, where

$$I_i = \int_{1/(2N)}^{1-1/(2N)} \{1 - x^n - (1-x)^n\}\, \phi_i(x)\, dx \qquad (1)$$

and p is the frequency of preferred codons in the sequence in question. With additive selection, we have [27]

$$\phi_0(x) = 2x^{-1}(1-x)^{-1}\{1 - \exp - \gamma(1-x)\}/\{1 - \exp - \gamma\}.$$

For $\phi_1(x)$, the sign of γ is reversed ([27], pp. 148).

p is given approximately by [4]

$$p = (\exp \gamma)/(k + \exp \gamma). \qquad (2)$$

Substituting this into equation (1), it is easily seen that the expression for π reduces to $(\exp \gamma)\, I_0/([\exp \gamma]\, I_0 + I_1)$, which is independent of k. This implies that γ can be estimated from the fraction of polymorphisms that are P→U. Let there be S segregating sites involving P→U or U→P mutations in a sample of n alleles, of which s are P→U mutations. On the assumption of binomial sampling of the proportion of P→U mutations, the log-likelihood of the sample is

$$\ln L = s \ln \pi + (S - s) \ln (1 - \pi). \qquad (3)$$

Data from different genes can be combined by summing the log-likelihood values for each gene. Given that the above correction for misclassified fixations assumes that mutations are neutral, it could introduce a bias in the presence of selection. Therefore, we used the uncorrected observed numbers of polymorphic changes for this analysis.

Estimate of the Rate of Recombination

The rate of recombination in the coding and intron regions of the gene *elav* was estimated using the method proposed by Wall [28]. The coding sequences were represented by the 50 alleles obtained here (Table 1) and the introns by a sample of 20 alleles of a 624 bp region of the adjacent intron 1 (kindly provided by J. Vieira). Although the exon and intron sequences have been obtained from different samples, this should not affect the analysis, since there is no evidence for genetic geographic differentiation at this locus along the area of distribution of this species (J. Vieira, personal communication).

Supplemental Data

Supplemental data including details on fly stocks, retrieval, and analysis of DNA sequences; two tables; and one figure are available at http://www.current-biology.com/cgi/content/full/14/2/150/DC1/.

Acknowledgments

We are grateful to H. Akashi and two anonymous reviewers for their comments, which greatly contributed to improving the manuscript, and also to G. Marais for helpful discussions and for retrieving *D. virilis* sequences from GeneBank. We are also indebted to C. Bartolomé for helping with the sequence collection, to L. Duret and P. Keightley for providing us with useful computer programs (see Experimental Procedures), to B. McAllister for sending us his *D. americana* collections, to J. Vieira and C. Vieira for sharing sequence data, and also to H. Cowan for helping with media preparation and washing up. B.C. is supported by the Royal Society of London and X.M. by the Biotechnology and Biological Sciences Research Council (BBSRC). This work was funded by a grant of the BBSRC to B.C.

Received: October 22, 2003
Revised: December 9, 2003
Accepted: December 10, 2003
Published: January 20, 2004

References

1. Li, W.H. (1987). Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. *24*, 337–345.
2. Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. Genetics *129*, 897–907.
3. Akashi, H., Kliman, R.M., and Eyre-Walker, A. (1998). Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. Genetica *103*, 49–60.
4. McVean, G., and Charlesworth, B. (1999). A population genetic model for the evolution of synonymnous codon usage: patterns and predictions. Genet. Res. *74*, 145–158.
5. Li, W.S. (1997). Molecular evolution (Sunderland, Massachusetts: Sinauer Associates, Inc.).
6. Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. *158*, 573–597.
7. Duret, L., and Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. Proc. Natl. Acad. Sci. USA *96*, 4482–4487.
8. Akashi, H. (1999). Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics *151*, 221–238.
9. Akashi, H. (1996). Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. Genetics *144*, 1297–1307.
10. McVean, G.A., and Vieira, J. (2001). Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in *Drosophila*. Genetics *157*, 245–257.
11. Machado, C.A., Kliman, R.M., Markert, J.A., and Hey, J. (2002). Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. *19*, 472–488.
12. Spicer, G.S., and Bell, C.D. (2002). Molecular phylogeny of the *Drosophila virilis* species group (Diptera: Drosophilidae) inferred from mitochondrial 12S and 16S ribosomal RNA genes. Ann. Entomol. Soc. Am. *95*, 156–161.
13. McAllister, B.F., and Charlesworth, B. (1999). Reduced sequence variability on the Neo-Y chromosome of *Drosophila americana americana*. Genetics *153*, 221–233.
14. McVean, G.A., and Vieira, J. (1999). The evolution of codon preferences in *Drosophila*: a maximum-likelihood approach to parameter estimation and hypothesis testing. J. Mol. Evol. *49*, 63–75.
15. Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. *2*, 13–34.
16. Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. Genetics *139*, 1067–1076.
17. Francino, M.P., and Ochman, H. (1999). Isochores result from mutation not selection. Nature *400*, 30–31.
18. Galtier, N., Piganeau, G., Mouchiroud, D., and Duret, L. (2001). GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. Genetics *159*, 907–911.
19. Marais, G. (2003). Biased gene conversion: implications for genome and sex evolution. Trends Genet. *19*, 330–338.
20. Marais, G., Mouchiroud, D., and Duret, L. (2001). Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc. Natl. Acad. Sci. USA *24*, 5688–5692.
21. Akashi, H. (1999). Within- and between-species DNA sequence variation and the 'footprint' of natural selection. Gene *238*, 39–51.
22. Fu, Y.X., and Li, W.H. (1993). Statistical tests of neutrality of mutations. Genetics *133*, 693–709.
23. Kliman, R.M., and Hey, J. (1994). The effects of mutation and natural selection on codon bias in the genes of *Drosophila*. Genetics *137*, 1049–1056.
24. Akashi, H., and Schaeffer, S.W. (1997). Natural selection and the frequency distributions of "silent" DNA polymorphism in *Drosophila*. Genetics *146*, 295–307.
25. Stenico, M., Lloyd, A.T., and Sharp, P.M. (1994). Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. Nucleic Acids Res. *22*, 2437–2446.
26. Penden, J.F. (1999). Analysis of codon usage. PhD thesis, University of Nottingham, UK.
27. Ewens, W. (1979). Mathematical population genetics (Berlin: Springer-Verlag).
28. Wall, J.D. (2000). A comparison of estimators of the population recombination rate. Mol. Biol. Evol. *17*, 156–163.